DOCUMENT RESUME

ED 090 826 HE 005 420

AUTHOR Turner, Richard L.; Thompson, Robert P.
TITLE Relationships Between College Student Ratings of Instructors and Residual Learning.
PUB DATE Sep 73
NOTE 6p.; Speech presented at the annual meetings of the American Educational Research Association, Chicago, Illinois, April 1974

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *College Students; *French; Graduate Students; *Higher Education; Research Projects; *Student Attitudes; *Teacher Evaluation

ABSTRACT

In 1971-72, 334 students in 16 sections in beginning French, and in 1972-73, 331 students in 24 sections in the same course, rated their graduate student instructors on a 35-item scale. Student performance data on the first, mid-term, and final departmental examinations and on SAT-V (1972-73 only) were taken, and residual learning gain (final exam corrected for first exam) computed. Cluster analysis yielded two intercorrelated clusters. Two scales with each item having unit weight were then developed as the student rating variables. A substantial negative correlation between the two rating scales and residual learning was replicated across the 2 years. (MJM)

US DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 090826

HE 005 420

20.10

Annual AERA Meeting
Chicago, April 1974

Relationships between College Student Ratings of Instructors and Residual Learning

Richard L. Turner and Robert P. Thompson
Indiana University[1]

September, 1973

Both student ratings of instructors and the amount students learn in a course, corrected for their initial learning (hence residual learning, or residual gain), are almost universally regarded as important criteria of instructor success. Of these two criteria, student ratings are the more easily obtained and the more widely used. Studies using student residual achievement as the criterion variable require the use of multiple section courses in which the same or very similar content is taught to students by different instructors. A uniform assessment of the students' knowledge or aptitude must be made prior to or early in the course and a uniform final assessment made at the end of the course, then the initial assessment regressed on the final assessment to obtain residual achievement scores. This procedure is time consuming and tedious and is rarely carried out. Typically the assumption is made that if students like or respond favorably to an instructor they learn well from him or her. It is this assumption which is tested in the studies reported here.

Procedures

Sample. In 1971-72, 334 students (99% of enrollees) in 16 sections of beginning French and in 1972-73, 331 students (63% of enrollees) in 24 sections in the same course, rated their graduate student instructors on a 35 item scale. Student performances on the first, mid-term and final department examinations and on the SAT-V (1972-73 only) were taken and residual learning gain (final exam corrected for first exam) computed.

Scale Development. Items representing each of the principal factors from the student rating scales developed by Deshpande, Webb and Marks (1970)[2] were discussed with the course director and subsequently taken to the French Department. Members of the department selected from each factor those items they believed most valid for instructors in their department, for a total of 30 items, and added five additional items specific to teaching beginning French. Following administration to the 1971-72 students in the final week of the course, the items were re-factored (principal components with varimax rotation) but an interpretable multi-factor solution could not be attained. A subsequent cluster analysis produced one large cluster and one small cluster closely correlated to it. Two subscales with each item having unit weight were then developed as the student rating variables. These scales were also used in the 1972-73 replication.

[1]The authors wish to acknowledge the extended cooperation of the Department of French and Italian and the support of the Research Committee on Teaching and Learning, Indiana University.

[2]Most or items in these scales are "positive" in wording. To avoid position habits among respondents, 12 items were worded negatively. Negatively worded items were reversed in score for the statistical analysis.

The median inter-observer reliability for random samples of six students from each section represented in the 1971-72 sample was, for the total scale, .93. The reliability (Cronbach's Alpha) of the total scale was .90 for subscale 1, .88 and for subscale 2, .64.

Test Development. Three teams, drawn from the instructors in the course, including the course director, made up the first, mid-term and final examinations, representing in the items the content covered in the course syllabus and in class. These examinations were administered on the same day by each instructor during the 3rd, 7th and last (15th) week of the course respectively. The first test covered grammar, and the remaining two grammar, dictation, composition and reading comprehension.

To develop the achievement criterion variable for the 1971-72 group, the first examination score was regressed on the final examination score ($r=.47$) using each student as one degree of freedom and a residual gain score for each student obtained. Subsequently, section means for the student rating scales, the first, mid-term and final examinations, and for residual gain, were calculated. These variables were then correlated with each other using each section as one degree of freedom. Identical procedures were used with the 1972-73 replication. In this replication the SAT-V scores were also correlated to the final exam ($r=.23$), to test it as a possible covariate. It was rejected as a useful covariate.

Results

The items appearing in each of the two subscales are shown in Table 1. The items in subscale 1 were originally associated with several first order factors extracted by Deshpande, Webb and Marks', including structure, rapport, motivation overload and content mastery. In the present analysis, however, these items seem to align with three of the four second order factors extracted by these authors, namely, cognitive merit, affective merit and stress. After inspecting the items, this subscale was thus labeled "Instructor cognitive and affective merit versus student cognitive and affective stress." Unlike subscale 1, subscale 2 appears to reflect two of Deshpande, Webb and Marks' first order factors, namely motivation and work overload, and is labeled in accordance.

The zero order correlations among the two student rating subscales, the first, mid-term and final examinations, and residual achievement gain for both 1971-72 and 1972-73 are shown in Table 2. In the 1971-72 study, each student rating subscale showed strong positive relationships to the first exam, with steady slippage across the later exams, culminating in significant negative correlations between the student ratings (positively interpreted) and residual achievement gain. A similar but less strong effect may be observed in the 1972-73 data.

Discussion

Examined from a research viewpoint, the results of the study appear to be congruent with the "validity studies" reported by McKeatchie, Lin and Mann (1971).

In these studies, the relationships between student ratings and residual achievement show both positive and negative signs, with the directionality of the relationship contingent upon the type of test given the sex of the student[3], and the particular factors in the rating scale. Among these variables, a relatively consistent relationship, for both sexes, is between "overload" as a factor in the rating scales and greater student performance on "knowledge" or factual type tests.

The same relationship appears to be present in the data reported here. The positive and negative items in subscale 1 seem to define a bipolar continuum of cognitive and affective merits of the instructor with stress being the negative pole. In cluster two, work overload appears as bipolar to positive motivation by the instructor as evidenced in teamwork, stimulus variation and inspiration to effort. Interpreted in accord with the bipolarity of the subscales, the results of the study suggest that the stress/overload produced by the instructor is the important factor in obtaining greater residual gain in beginning French. Collaterally, the meritorious or positive behaviors of the instructor, those which make the course more enjoyable, manageable and orderly from a student viewpoint, appear to lead to less residual gain. Although results of this type are not very palatable, their congruence with the McKeatchie data suggest that teacher behaviors which lead to student stress and overload be given careful attention in subsequent research.

Examined from a practical viewpoint, the results of the study strongly suggest that student ratings of college instructors should be treated with great caution by college administrators and by promotion and tenure committees. Although such ratings may express student observations of and attitudes toward an instructor, they clearly cannot be routinely interpreted to be positive indicators of student residual achievement in the instructor's course.

References

Deshpande, A.S., Webb, S.C. and Marks, E. "Student Perceptions of Engineering Instructor Behaviors and Their Relationships to the Evaluation of Instructors and Courses." AERJ, VII, 289-305, May 1970.

McKeatchie, W.J., Lin, Y.G., and Mann, W. "Student Ratings of Teacher Effectiveness: Validity Studies." AERJ, VIII, 435-445, May, 1971.

[3]The small section size in the present study prevented a reliable analysis for sex effects.

Table 1. Student Rating Item-Subscale Correlations

Subscale 1: Instructor cognitive and affective merit versus student cognitive and affective stress.	r with subscale
Positive Items	
41. Seemed concerned that students learn	.66
23. Put the subject across in a lively way	.65
47. Explained clearly and his explanations were to the point	.64
29. Made you want to do your best in the course	.60
28. Gave students frequent opportunities to speak in French	.57
29. Provided appropriate correction and guidance in spoken work	.56
30. Had a good command of French	.56
40. Seemed to have a thorough knowledge of cultures of French speaking peoples	.55
18. Was well prepared each day	.54
42. Arranged his presentation logically	.50
16. Kept course moving at an even, steady pace	.50
33. Was friendly	.46
30. Seemed sure of himself in front of the class	.45
27. Showed concern for students as persons	.41
49. Created an atmosphere in which students in the class seemed friendly	.38
19. Provided time for questions and discussion	.38
25. Expressed concepts at a level understandable by students	.24
Negative Items	
40. Gave vague explanations	.67
22. Made students feel afraid of him/her	.66
32. Seemed disorganized	.63
36. "Talked down" to students	.60
20. Made frequent pronunciation errors in French	.60
24. Seemed confused in what he was doing	.56
17. Pitched his presentations above the heads of students	.56
31. Did not seem to like or understand students	.55
15. Could not explain text materials that were confusing to students	.52
37. Made the course unnecessarily difficult	.49
34. Showed little enthusiasm for teaching French	.31

Alpha=.88

Table 1, continued

Subscale 2: Workload and motivation	r with subscale
Positive Items	
43. Required a reasonable amount of work	.64
39. Inspired you to independent efforts	.63
44. Provided a variety of activities in class and used a variety of media (slides, films, projections, drawings) and outside resource persons	.62
45. Encouraged class members to work as a team	.56
35. Avoided assigning a lot of burdensome busywork	.49
Negative Item	
46. Asked more than students could get done	.67

Alpha=.64

Table 2. Intercorrelations among Student Ratings, Examinations and Residual Achievement

1971-72 n=16

	Subscale 2	Subscale 1+2	1st Exam	Midterm	Final Exam	Residual
Subscale 1	90**	99**	73**	08	-15	-51*
Subscale 2		94**	6[illegible]**	-07	-28	-52*
Subscale 1+2			71**	05	-18	-52*
First Exam				18	13	-49*
Midterm					77**	60*
Final Exam						77**

1972-73 n=24

	Subscale 2	Subscale 1+2	SATV	1st Exam	Midterm	Final Exam	Residual
Subscale 1	73**	98**	06	22	30	08	-41*
Subscale 2		85**	-04	25	05	12	-31
Subscale 1+2			04	24	25	09	-41*
SATV				-06	-01	02	09
1st Exam					53**	66**	-14
Midterm						41*	-02
Final							58**

* P .05
** P .01